



The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the β 7- β 8 and β 9- β 10 hairpin loops and conserved zinc binding domain cysteines

Tatjana P. Kristensen^{1†‡}, Reeja Maria Cherian^{1†‡}, Fiona C. Gray¹ and Stuart A. MacNeill^{1,2*}

¹ Department of Biology, University of Copenhagen, Københavns Biocenter, Copenhagen N, Denmark

² School of Biology, University of St. Andrews, North Haugh, St. Andrews, Fife, UK

Edited by:

R. Thane Papke, University of Connecticut, USA

Reviewed by:

Nils-Kåre Birkeland, University of Bergen, Norway

Antonio Ventosa, University of Sevilla, Spain

*Correspondence:

Stuart A. MacNeill, School of Biology, University of St. Andrews, North Haugh, St. Andrews, Fife, KY16 9ST, UK
e-mail: stuart.macneill@st-andrews.ac.uk

† Present address:

Tatjana P. Kristensen, Department of Veterinary Disease Biology, University of Copenhagen, Frederiksberg C, Denmark;
Reeja Maria Cherian, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden

‡ These authors have contributed equally to this work.

The hexameric MCM complex is the catalytic core of the replicative helicase in eukaryotic and archaeal cells. Here we describe the first *in vivo* analysis of archaeal MCM protein structure and function relationships using the genetically tractable haloarchaeon *Haloferax volcanii* as a model system. *Hfx. volcanii* encodes a single MCM protein that is part of the previously identified core group of haloarchaeal MCM proteins. Three structural features of the N-terminal domain of the *Hfx. volcanii* MCM protein were targeted for mutagenesis: the β 7- β 8 and β 9- β 10 β -hairpin loops and putative zinc binding domain. Five strains carrying single point mutations in the β 7- β 8 β -hairpin loop were constructed, none of which displayed impaired cell growth under normal conditions or when treated with the DNA damaging agent mitomycin C. However, short sequence deletions within the β 7- β 8 β -hairpin were not tolerated and neither was replacement of the highly conserved residue glutamate 187 with alanine. Six strains carrying paired alanine substitutions within the β 9- β 10 β -hairpin loop were constructed, leading to the conclusion that no individual amino acid within that hairpin loop is absolutely required for MCM function, although one of the mutant strains displays greatly enhanced sensitivity to mitomycin C. Deletions of two or four amino acids from the β 9- β 10 β -hairpin were tolerated but mutants carrying larger deletions were inviable. Similarly, it was not possible to construct mutants in which any of the conserved zinc binding cysteines was replaced with alanine, underlining the likely importance of zinc binding for MCM function. The results of these studies demonstrate the feasibility of using *Hfx. volcanii* as a model system for reverse genetic analysis of archaeal MCM protein function and provide important confirmation of the *in vivo* importance of conserved structural features identified by previous bioinformatic, biochemical and structural studies.

Keywords: *Haloferax volcanii*, archaea, Haloarchaea, MCM helicase, DNA replication, reverse genetics, zinc binding domain

INTRODUCTION

In all forms of life, successful chromosomal DNA replication requires efficient unwinding of the DNA double helix at the replication forks, a reaction catalyzed by the replicative DNA helicase. In eukaryotes the replicative helicase is the CMG complex, a tripartite molecular machine composed of Cdc45, MCM and GINS (reviewed by Onesti and MacNeill, 2013). The MCM (mini-chromosome maintenance) complex is the catalytic core of this machine (Vijayraghavan and Schwacha, 2012). MCM is a ring-shaped hexamer composed of six related but non-identical subunits, each of which is a member of the AAA+ (ATPases associated with diverse cellular activities) protein superfamily (Duderstadt and Berger, 2008). MCM is loaded onto chromosomal replication origins as a head-to-head double hexamer in the G1 phase of the cell cycle in a reaction known as replication licensing (Evrin et al., 2009; Remus et al., 2009; Gambus et al., 2011). Cdc45 and GINS then assemble at the G1-S boundary to form the CMG, activation of which involves MCM subunit

phosphorylation by DDK (Dbf4-dependent protein kinase). Once activated, individual CMG complexes move with the replication forks from origin to inter-origin sequences.

Consistent with their shared evolutionary history, homologs of the major eukaryotic replication factors have been identified and characterized in the archaea, including homologs of the three components of the CMG. Owing to their high sequence similarity to their eukaryotic counterparts, archaeal MCM proteins were the first to be identified and biochemically characterized. Many archaea encode single MCM proteins that have been shown to form—or are presumed to form—homohexameric helicase complexes (reviewed by Slaymaker and Chen, 2012). The best studied examples of this type of MCM are from the euryarchaeon *Methanothermobacter thermoautotrophicum* and the crenarchaeon *Sulfolobus solfataricus*. However, a number of species encode multiple MCM proteins, such as *Thermococcus kodakarensis* and *Methanococcus maripaludis*, which encode three and four, respectively (Walters and Chong, 2010; Ishino et al.,

2011; Pan et al., 2011). It is not impossible that in some species these proteins could form eukaryotic-like heterohexameric complexes *in vivo*. Interestingly, only one of the three *T. kodakarensis* MCM proteins is essential for cell viability (Ishino et al., 2011; Pan et al., 2011).

Structural information is available for a number of archaeal MCM proteins and has been used to guide biochemical investigations of protein structure-function relationships (reviewed by Slaymaker and Chen, 2012). At a structural level, individual archaeal MCM proteins comprise a non-catalytic N-terminal domain followed by the catalytic AAA+ domain and, at the extreme C-terminus, a short winged helix-turned-helix (wHTH) domain. The most extensive crystal structure is that of near full-length *S. solfataricus* MCM which spans the N-terminal and catalytic AAA+ domains but not the wHTH domain (Brewster et al., 2008). Efforts to determine the structure of the wHTH by NMR are ongoing (Wiedemann et al., 2013). Additional crystal structures include the N-terminal domains of *S. solfataricus*, *M. thermoautotrophicum*, and *Thermoplasma acidophilum* MCM proteins, with the latter forming a right-handed spiral filament (Fletcher et al., 2003; Liu et al., 2008; Fu et al., 2014). A left-handed filament structure for near full-length *S. solfataricus* MCM has also been determined (Slaymaker et al., 2013), as well as a full-length structure of a catalytically inactive MCM homolog from *Methanopyrus kandleri* (Bae et al., 2009). The biological significance, if any, of the filamentous forms remains to be determined.

Unlike the MCM proteins, archaeal GINS and Cdc45 homologs share only very limited sequence similarity with their eukaryotic counterparts (Marinsek et al., 2006). The eukaryotic GINS complex is a heterotetramer, comprising the related Sld5, Psf1, Psf2, and Psf3 proteins (reviewed by Kamada, 2012). Both homotetrameric and heterotetrameric (i.e., dimer of dimer or A₂B₂) complexes have been identified in archaea and the structure of the *T. kodakarensis* A₂B₂ heterotetrameric GINS has been solved (Oyama et al., 2011). Archaeal Cdc45 homologs have only very recently been positively identified as such (Sanchez-Pulido and Ponting, 2011; Krastanova et al., 2012; Makarova et al., 2012). These proteins belong to the RecJ nuclease branch of the DHH hydrolase superfamily. Unlike eukaryotic Cdc45 proteins, at least some archaeal RecJ/Cdc45 proteins possess nuclease activity (Li et al., 2011; Yuan et al., 2013), the precise function of which is unclear. The existence of all three CMG components in archaea suggest that these organisms may have a valuable role to play as models for dissecting the function of the individual CMG components.

Using multiple sequence alignments and crystal structures as a guide, a number of laboratories have reported detailed mutagenesis studies of MCM structure-function relationships (reviewed by Slaymaker and Chen, 2012). In all cases, the effects of the mutations on MCM function were determined *in vitro*, using purified recombinant proteins in various biochemical assays. To our knowledge, there has been no *in vivo* reverse genetic analysis of the effects of mutations of MCM function, largely due to the difficulty or impossibility of conducting such studies in species where genetic tools are either

rudimentary or unavailable. In this report we describe the first results of reverse genetic analysis of archaeal MCM function *in vivo*, using the haloarchaeal organism *Haloferax volcanii* as a model system. The haloarchaea present a particularly attractive model to study archaeal chromosome replication owing to the ease with which representative species can be manipulated genetically (reviewed by Farkas et al., 2013). *Hfx. volcanii* in particular has proved a highly successful model, with a number of components of the *Hfx. volcanii* replication machinery already characterized, including multiple origins of replication (Hawkins et al., 2013), origin binding proteins (Norais et al., 2007), single-stranded DNA binding proteins (Skowrya and MacNeill, 2012; Stroud et al., 2012), the sliding clamp PCNA (Morgunova et al., 2009; Winter et al., 2009) and both ATP- and NAD-dependent DNA ligases (Poidevin and MacNeill, 2006; Zhao et al., 2006). In addition to their genetic tractability, over 100 haloarchaeal genomes have been now sequenced, offering a wealth of information for comparative protein sequence analysis. *Hfx. volcanii* encodes a single MCM protein, a member of the previously defined core group of haloarchaeal MCM proteins discussed further below (MacNeill, 2009). In the work presented here, three conserved features of the protein are targeted for mutagenesis: the β 7- β 8 and β 9- β 10 β -hairpin loops and the four conserved cysteines of the putative zinc binding domain. The results of these studies establish *Hfx. volcanii* as a valuable model for detailed structure-function analysis of MCM helicase and provide confirmation of the importance of these conserved features for MCM function *in vivo*.

MATERIALS AND METHODS

DATABASE SEARCHING AND SEQUENCE HANDLING

Protein sequences were obtained from the UniProt Knowledgebase (UniProtKB) database (Magrane and UniProt Consortium, 2011); primary accession numbers are listed in **Table 1**. Multiple sequence alignments and phylogenetic trees were generated using ClustalX 2.1 (Larkin et al., 2007) and njplot (Perriere and Gouy, 1996), respectively. Intein sequences were initially identified by visual inspection as large sequence insertions in comparative sequence analysis. Inteins boundaries were defined by the presence of the N-terminal (block A) and C-terminal (block G) intein splicing motifs as defined at the InBase intein database (Perler, 2002).

STRAINS AND GROWTH CONDITIONS

Hfx. volcanii strains used in this study are listed in **Table 2**. All strains were grown in either Hv-YPC or Hv-CA medium at 45°C as described in the Halohandbook v7.2 (www.haloarchaea.com/resources/halohandbook). For selection procedures, tryptophan was added to Hv-CA medium at a final concentration of 50 μ g/ml. For counter-selection using 5-fluoroorotic acid (5-FOA), Hv-CA was supplemented with uracil and 5-FOA at final concentrations of 10 μ g/ml and 50 μ g/ml, respectively. For mitomycin C sensitivity assays, wild-type (H26) and mutant cells were grown at 45°C in Hv-YPC medium to an OD_{650 nm} of 0.2–0.32, before being serially diluted in 18% SW. 5 μ l aliquots were then spotted on Hv-YPC plates containing 0, 10, 20, or 30 ng/ml mitomycin C and incubated at 45°C for 5 days.

Table 1 | List of MCM proteins analyzed in this study.

Key to Figure 1	Species	UniProtKB accession number	Length (amino acids)	Inteins			
				A	B	C	D
1	<i>Haladaptatus paucihalophilus</i> DX253	E7QNU9	698				
2	<i>Halalkalicoccus jeotgali</i> B3	D8J3U5	700				
3A	<i>Haloarcula marismortui</i> ATCC 43049	Q5UYX8	1175	+			
3B		Q5V011	681				
3C		Q5V814	649				
4	<i>Halobacterium salinarum</i> R1	B0R796	879	+			
5A	<i>Halobiforma lacisalsi</i> AJ5	M0LR18	1342			+	
5B		M0LZ47	312				
6	<i>Halococcus thailandensis</i> JCM 13552	M0N8X7	698				
7	<i>Haloferax volcanii</i> DS2	D4GZG5	702				
8	<i>Halogeometricum borinquense</i> DSM 11551	E4NRK9	1818	+		+	
9	<i>Halogramum salarium</i> B-1	J2ZGA0	700				
10A	<i>Halomicrobium mukohataei</i> DSM 12286	C7P1F9	873	+			
10B		C7NZ81	458				
11A	<i>Halopiger xanaduensis</i> SH-6	F8D3Z0	702				
11B		F8DEM3	698				
11C		F8DET6	315				
12	<i>Haloquadratum walsbyi</i> HBSQ001	Q18E84	2216	+	+	+	+
13	<i>Halorhabdus utahensis</i> AX-2	C7NUH7	1412	+		+	
14A	<i>Halorubrum lacusprofundi</i> DSM 5036	B9LTB1	700				
14B		B9LUI3	717				
15	<i>Halosarcina pallida</i> JCM 14848	M0D6S0	1172	+			
16A	<i>Halosimplex carlsbadense</i> 2-9-1	M0D2C8	698				
16B		M0CE93	712				
17A	<i>Haloterrigena turkmenica</i> DSM 5511	D2RUS4	700				
17B		D2S3H9	314				
18	<i>Halovivax asiaticus</i> JCM 14624	M0BKP0	876	+			
19A	<i>Natrialba asiatica</i> DSM 12278	M0AKM2	700				
19B		M0AIF1	315				
20	<i>Natrinema pallidum</i> DSM 3751	L9Z845	700				
21A	<i>Natronobacterium gregoryi</i> SP2	L0AHL1	1172	+			
21B		L0AN78	709				
21C		L0ANL5	662				
22	<i>Natronococcus amylolyticus</i> DSM 10524	L9XCC3	1404	+	+		
23	<i>Natronolimnobius innermongolicus</i> JCM 12255	L9X2E8	700				
24A	<i>Natronomonas pharaonis</i> DSM 2160	Q3IML4	1037	+			
24B		Q3IPB6	676				
25	<i>Natronorubrum tibetense</i> GA33	L9VHE4	1814	+		+	
26	<i>Salinarchaeum</i> sp. Harcht-Bsk1	R4W989	697				
	<i>Methanosarcina acetivorans</i> C2A	Q8TSW4	701				
	<i>Methanothermobacter thermautotrophicus</i> ΔH	Q27798	666				
	<i>Sulfolobus solfataricus</i> P2	Q9UXG1	686				
	<i>Cenarchaeum symbiosum</i> A	A0RYB8	697				
	<i>Korarchaeum cryptofilum</i> OPF8	B1L6L9	703				
	<i>Thermoplasma acidophilum</i> DSM 1728	Q9HK10	698				
	<i>Archaeoglobus fulgidus</i> DSM 4304	Q29733	586				
	<i>Homo sapiens</i> Mcm2	P49736	904				
	<i>Homo sapiens</i> Mcm3	P25205	808				
	<i>Homo sapiens</i> Mcm4	P33991	863				
	<i>Homo sapiens</i> Mcm5	P33992	734				
	<i>Homo sapiens</i> Mcm6	Q14566	821				
	<i>Homo sapiens</i> Mcm7	P33993	719				

The numbers shown in the first column provide a key for the proteins represented in the phylogenetic tree shown in **Figure 1**. The plus signs in the final four columns indicate the present of intein insertions at each of the four conserved positions A–D shown in **Figure 2**. Protein 3A, for example, Q5UYX8 from *Haloarcula marismortui*, has a single intein at position A.

Table 2 | *Haloferax volcanii* strains used in this study.

Strain no.	Genotype	Notes	References
SMH693	–	Wild-type strain DS70	Wendoloski et al., 2001
SMH630	$\Delta pyrE2$	Strain H26	Allers et al., 2004
SMH656	<i>mcm-rM1</i> $\Delta pyrE2$	Single alanine substitution Q186A	This study
SMH654	<i>mcm-rM3</i> $\Delta pyrE2$	Single alanine substitution E190A	This study
SMH658	<i>mcm-rM4</i> $\Delta pyrE2$	Single alanine substitution R193A	This study
SMH660	<i>mcm-rM5</i> $\Delta pyrE2$	Single alanine substitution E196A	This study
SMH662	<i>mcm-rM6</i> $\Delta pyrE2$	Single alanine substitution Q199A	This study
SMH638	<i>mcm-bH1</i> $\Delta pyrE2$	Paired alanine substitution H226A/I227A	This study
SMH640	<i>mcm-bH2</i> $\Delta pyrE2$	Paired alanine substitution E228A/Q229A	This study
SMH642	<i>mcm-bH3</i> $\Delta pyrE2$	Paired alanine substitution Q230A/T231A	This study
SMH644	<i>mcm-bH4</i> $\Delta pyrE2$	Paired alanine substitution S232A/G233A	This study
SMH646	<i>mcm-bH5</i> $\Delta pyrE2$	Paired alanine substitution N234A/E235A	This study
SMH648	<i>mcm-bH6</i> $\Delta pyrE2$	Paired alanine substitution K236A/T237A	This study
SMH649	<i>mcm-hD1</i> $\Delta pyrE2$	Two amino acid deletion T231/S232	This study
SMH650	<i>mcm-hD2</i> $\Delta pyrE2$	Four amino acid deletion Q230/T231/S232/G233	This study
SMH652	<i>mcm-S1</i> $\Delta pyrE2$	Silent <i>Acc65I</i> restriction site replacing codons 141 and 142	This study

MOLECULAR CLONING REAGENTS

Enzymes for molecular cloning were purchased from New England Biolabs (NEB), Promega or Fermentas. Oligonucleotides were synthesized by DNA Technology A/S (Risskov, Denmark). DNA sequencing was performed by Eurofins MWG Operon (Ebersberg, Germany). DNA purification kits were from Qiagen. PCR amplification was performed using the GC-rich PCR system (Roche) with *Taq* polymerase (NEB) substituting for the GC-rich enzyme as necessary. For routine cloning purposes, *E. coli* DH5 α (*fhuA2* Δ (*argF-lacZ*) *U169 phoA glnV44* Φ 80 Δ (*lacZ*) *M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used (Invitrogen). To prepare unmethylated plasmid DNA for *Hfx. volcanii* transformation, *E. coli* GM121 (F- *dam-3 dcm-6 ara-14 fhuA31 galK2 galT22 hdsR3 lacY1 leu-6 thi-1 thr-1 tsx-78*) was used.

CONSTRUCTION OF MUTANT *Hfx. Volcanii* STRAINS

Mutant strains were constructed using the pop-in/pop-out method (Bitan-Banin et al., 2003) in *Hfx. volcanii* $\Delta pyrE2$ strain H26 as follows. First, plasmid pTA131-HfxMCM-HXba was constructed by using oligonucleotide primers HfxMCM-5H and HfxMCM-3X (designed to include *HindIII* and *XbaI* sites respectively, see Table 3) to amplify a 1.0 kb fragment of *Hfx. volcanii* genomic DNA spanning the region from 100 nucleotides upstream of the *mcm* ORF to 900 nucleotides inside the ORF. The PCR product was digested with *HindIII* and *XbaI*, cloned into plasmid pTA131 digested with the same two enzymes (Allers et al., 2004) and sequenced to confirm the absence of PCR errors.

The resulting plasmid (pTA131-HfxMCM-HXba) was then used as a template for PCR overlap extension mutagenesis (OEM) to create 25 mutant derivatives. The sequences of the mutagenic primers are given in Table 3. OEM was performed using the GC-Rich PCR system buffer and GC-Rich enzyme mix (Roche) for the first round of PCR and the GC-Rich PCR system buffer and *Taq* polymerase (NEB) for the second round. Oligonucleotides HfxMCM-5H and HfxMCM-3X were used as

the flanking primers throughout. The resulting PCR products were then restricted with *Bsu36I* and *BstEII* (for β 7- β 8 β -hairpin and zinc binding domain mutants) or *BstEII* and *XbaI* (for β 9- β 10 β -hairpin mutants) and the 329 bp *Bsu36I*-*BstEII* or 271 bp *BstEII*-*XbaI* pieces carrying the mutations cloned back into pTA131-HfxMCM-HXba from which the corresponding wild-type restriction fragments had been removed. Plasmids were again sequenced to confirm the absence of unwanted sequence changes before being passaged through *E. coli* GM121 to generate unmethylated DNA for transformation into *Hfx. volcanii* $\Delta pyrE2$ strain H26 (Table 2).

Transformation of *Hfx. volcanii* was accomplished as described in the Halohandbook v7.2 (www.haloarchaea.com/resources/halohandbook). Transformants obtained on Hv-CA medium lacking uracil were grown for 30 generations at 45°C in non-selective Hv-YPC medium to allow loss of the plasmid before being plated on Hv-CA plates containing 50 μ g/ml 5-FOA and a 10 μ g/ml uracil. Colonies formed on these plates were then inoculated into 500 μ l of Hv-YPC liquid medium and grown overnight at 45°C. Genomic DNA from these was prepared by taking 10 μ l of overnight culture, adding to 500 μ l of sterile water and heating to 70°C for 10 min to lyse the cells. For the β -hairpin and zinc binding domain mutants, 1 μ l of the resulting mix was used in PCR reactions to screen for the presence of the mutants in the chromosome. To discriminate between wild-type and mutant sequences, oligonucleotide primers with mismatched 3' sequences were used in PCR reactions in conjunction with primer HFXMCM-R1150, which lies 1150 nucleotides into the *mcm* ORF and which is therefore not present in pTA131-HfxMCM-HXba (see Table 3 for oligonucleotide sequences). For the silent restriction site mutant, the partial *mcm* ORF was amplified using primers HfxMCM-5H and HfxMCM-3X and the PCR products digested with *Acc65I* to identify mutants. In all cases, putative positive clones were then re-streaked twice to single colonies on Hv-YPC agar at

Table 3 | Oligonucleotides used in this study.

Primer name	Sequence
A. FOR CONSTRUCTION OF pTA131-HfxMCM-HXba	
HfxMCM-5H	5'-GTGTGTGTGTAAGCTTCTCCGCGAGGCGACGGA-3'
HfxMCM-3X	5'-GGTGGTGGTGCTAGACATGGCAATCTTCTCTG-3'
B. FOR CONSTRUCTION OF β7-β8 β-HAIRPIN MUTANTS BY OEM	
HFXMCM-rM1-F	5'-AAACTGCGCGTCGCGGAGTCCCCGAGGGC-3'
HFXMCM-rM1-R	5'-GCCCTCGGGGACTCGGCGACGCGCAGTTT-3'
HFXMCM-rM2-F	5'-AAACTGCGCGTCCAGGCTCCCCGAGGGCCTGCG-3'
HFXMCM-rM2-R	5'-CGCAGGCCCTCGGGGAGGCGTGGACGCGCAGTTT-3'
HFXMCM-rM3-F	5'-CAGGAGTCCCCGCGGCGCTGCGCGGGGCG-3'
HFXMCM-rM3-R	5'-GCCCGCGCGCAGGCGGCGGGGACTCCTG-3'
HFXMCM-rM4-F	5'-CCCGAGGGCGCTGGCGGGGGGCGAGACGCCG-3'
HFXMCM-rM4-R	5'-CGGCGTCTCGCCCCCGCAGGCCCTCGGG-3'
HFXMCM-rM5-F	5'-CTGCGCGGGGGCGGACGCCGCGAGCATC-3'
HFXMCM-rM5-R	5'-GATGCTCTGCGGCGTGCGGCCCGCGCAG-3'
HFXMCM-rM6-F	5'-GGCGAGACGCCGCGCAGCATCGACATCAAC-3'
HFXMCM-rM6-R	5'-GTTGATGTCGATGCTGGCGGCGTCTCGCC-3'
HFXMCM-rD1-F	5'-CAGGAGTCCCCGAG-GGCGAGACGCCGCGAG-3'
HFXMCM-rD1-R	5'-CTGCGGCGTCTCGCC-CTCGGGGACTCCTG-3'
HFXMCM-rD2-F	5'-GTCCAGGAGTCCCC-GAGACGCCGCGAGC-3'
HFXMCM-rD2-R	5'-GCTCTGCGGCGTCTC-GGGGGACTCCTGGAC-3'
HFXMCM-rD3-F	5'-CGCGTCCAGGAGTCC-ACGCCGCGAGCATC-3'
HFXMCM-rD3-R	5'-GATGCTCTGCGGCGTGACTCCTGGACGCG-3'
HFXMCM-rD4-F	5'-CTGCGCGTCCAGGAG-CCGCGAGCATCGAC-3'
HFXMCM-rD4-R	5'-GTCGATGCTCTGCGG-CTCCTGGACGCGCAG-3'
C. FOR CONSTRUCTION OF β9-β10 β-HAIRPIN MUTANTS BY OEM	
HFXMCM-bH1-F	5'-GTCGCGCTCCTCGCAGCGGAACAGCAGACATCG-3'
HFXMCM-bH1-R	5'-CGATGTCTGCTGTTCGCTGCGAGGACGCCGAC-3'
HFXMCM-bH2-F	5'-GTCCTCCACATCGCAGCGCAGACATCGGGCAAC-3'
HFXMCM-bH2-R	5'-GTTGCCCGATGTCTGCGCTGCGATGTGGAGGAC-3'
HFXMCM-bH3-F	5'-CACATCGAACAGCGCAGCGTCCGGCAACGAGAAG-3'
HFXMCM-bH3-R	5'-CTTCTCGTTGCCGACGCTGCTGTTCGATGTG-3'
HFXMCM-bH4-F	5'-GAACAGCAGACAGCAGCGAACGAGAAGACGCC-3'
HFXMCM-bH4-R	5'-GGGCGTCTTCTCGTTGCTGCTGTCTGCTTC-3'
HFXMCM-bH5-F	5'-CAGACATCGGGCGCAGCGAAGACGCCGCTCTTC-3'
HFXMCM-bH5-R	5'-GAAGACGGGCGTCTTCGCTGCGCCGATGTCTG-3'
HFXMCM-bH6-F	5'-TCGGGCAACGAGGCGCGCCGCTCTCGACTAC-3'
HFXMCM-bH6-R	5'-GTAGTCGAAGACGGGCGTCCCTGTTGCCGCA-3'
HFXMCM-hD1-F	5'-CACATCGAACAGCAG-GGCAACGAGAAGACG-3'
HFXMCM-hD1-R	5'-CGTCTTCTCGTTGCC-CTGCTGTTGATGTG-3'
HFXMCM-hD2-F	5'-CTCCACATCGAACAG-AACGAGAAGACGCC-3'
HFXMCM-hD2-R	5'-GGGCGTCTTCTCGTT-CTGTTGATGTGGAG-3'
HFXMCM-hD3-F	5'-GTCCTCCACATCGAA-GAGAAGACGCCGTC-3'
HFXMCM-hD3-R	5'-GACGGGCGTCTTCTC-TTCGATGTGGAGGAC-3'
HFXMCM-hD4-F	5'-GGGCGTCTCCACATC-AAGACGCCGCTTTC-3'
HFXMCM-hD4-R	5'-GAAGACGGGCGTCTT-GATGTGGAGGACGCC-3'
D. FOR CONSTRUCTION OF ZINC BINDING DOMAIN MUTANTS BY OEM	
HFXMCM-C137A-F	5'-CCGCCTTCGAGGCGCAGCGCTGCGG-3'
HFXMCM-C137A-R	5'-CCGACAGCGTGCCTCGAAGGCGG-3'
HFXMCM-C140A-F	5'-AGTGCCAGCGCGCGGGACGATGAG-3'
HFXMCM-C140A-R	5'-CTCATCGTCCCGCGCGCTGGCACT-3'
HFXMCM-C159A-F	5'-AACCCACAGAGCGCAGGGATGCGA-3' 5'-
HFXMCM-C159A-R	TCGCATCCCTGCGCCTCGTGGGGTT-3'
HFXMCM-C162A-F	5'-AGTGTGAGGAGCGGAGCGCCAGGG-3'
HFXMCM-C162A-R	5'-CCCTGGCGCTCCGCTCCCTGACACT-3'
E. FOR CONSTRUCTION OF SILENT Acc65I mutation by OEM	
HFXMCM-S1-F	5'-TGCCAGCGTGCCTACCATGAGTACATC-3'
HFXMCM-S1-R	5'-GATGTAGTCTATGTTACCGCAGCGCTGGCA-3'

(Continued)

Table 3 | Continued

Primer name	Sequence
F. FOR DETECTION OF β7-β8 β-HAIRPIN MUTANTS	
HFXMCM-rM1-WT	5'-CAGAACTGCGCGTCCAG-3'
HFXMCM-rM1-MUT	5'-CAGAACTGCGCGTCCGCG-3'
HFXMCM-rM2-WT	5'-AAACTGCGCGTCCAGGAG-3'
HFXMCM-rM2-MUT	5'-AAACTGCGCGTCCAGGCC-3'
HFXMCM-rM3-WT	5'-GTCCAGGAGTCCCCGAG-3'
HFXMCM-rM3-MUT	5'-GTCCAGGAGTCCCCCGCG-3'
HFXMCM-rM4-WT	5'-TCCCCCGAGGGCCTGCGC-3'
HFXMCM-rM4-MUT	5'-TCCCCCGAGGGCCTGGCG-3'
HFXMCM-rM5-WT	5'-GGCCTGCGCGGGGGCGAG-3'
HFXMCM-rM5-MUT	5'-GGCCTGCGCGGGGGCGCG-3'
HFXMCM-rM6-WT	5'-GGGGGCGAGACGCCGCG-3'
HFXMCM-rM6-MUT	5'-GGGGGCGAGACGCCGCC-3'
HFXMCM-rD1-MUT	5'-GGAGTCCCCGAGGGCGAG-3'
HFXMCM-rD2-MUT	5'-CCAGGAGTCCCCGAGAC-3'
HFXMCM-rD3-MUT	5'-CGTCCAGGAGTCCACGCC-3'
HFXMCM-rD4-MUT	5'-CTGCGCGTCCAGGAGCCG-3'
G. FOR DETECTION OF β9-β10 β-HAIRPIN MUTANTS	
HFXMCM-bH1-WT	5'-GTCGCGCTCCTCCACATC-3'
HFXMCM-bH1-MUT	5'-GTCGCGCTCCTCGCAGCG-3'
HFXMCM-bH2-WT	5'-GTCCTCCACATCGAACAG-3'
HFXMCM-bH2-MUT	5'-GTCCTCCACATCGCAGCG-3'
HFXMCM-bH3-WT	5'-CACATCGAACAGCAGACA-3'
HFXMCM-bH3-MUT	5'-CACATCGAACAGGCGAGCG-3'
HFXMCM-bH4-WT	5'-GAACAGCAGACATCGGGC-3'
HFXMCM-bH4-MUT	5'-GAACAGCAGACAGCAGCG-3'
HFXMCM-bH5-WT	5'-CAGACATCGGGCAACGAG-3'
HFXMCM-bH5-MUT	5'-CAGACATCGGGCGCAGCG-3'
HFXMCM-bH6-WT	5'-ACATCGGGCAACGAGAAGA-3'
HFXMCM-bH6-MUT	5'-ACATCGGGCAACGAGGCGAG-3'
HFXMCM-hD1-WT	5'-CACATCGAACAGCAGACA-3'
HFXMCM-hD1-MUT	5'-CACATCGAACAGGCGCG-3'
HFXMCM-hD2-WT	5'-CTCCACATCGAACAGCAG-3'
HFXMCM-hD2-MUT	5'-CTCCACATCGAACAGAAC-3'
HFXMCM-hD3-WT	5'-TCCTCCACATCGAACAGC-3'
HFXMCM-hD3-MUT	5'-TCCTCCACATCGAAGAGA-3'
HFXMCM-hD4-WT	5'-GGCGTCTCCACATCGAA-3'
HFXMCM-hD4-MUT	5'-GGCGTCTCCACATCAAG-3'
H. FOR DETECTION OF ZINC BINDING DOMAIN MUTANTS	
HFXMCM-C137A-WT	5'-GAAGCCGCTTCGAGTGC-3'
HFXMCM-C137A-MUT	5'-GAAGCCGCTTCGAGGCG-3'
HFXMCM-C140A-WT	5'-TTCGAGTGCCAGCGCTGC-3'
HFXMCM-C140A-MUT	5'-TTCGAGTGCCAGCGCGCG-3'
HFXMCM-C159A-WT	5'-CAGGAACCCACGAGTGT-3'
HFXMCM-C159A-MUT	5'-CAGGAACCCACGAGGCG-3'
HFXMCM-C162A-WT	5'-CACGAGTGTGAGGATGC-3'
HFXMCM-C162A-MUT	5'-CACGAGTGTGAGGAGCG-3'
I. REVERSE PRIMER FOR MUTANT DETECTION	
HFXMCM-R1150	5'-GCGAATCCGCGAGCCGTC-3'

Restriction sites in oligonucleotide primers used for pTA131-HfxMCM-HXba construction underlined. Oligonucleotide primers for overlap extension mutagenesis (OEM) are shown in top strand-bottom strand pairs with the mutated bases underlined in the top strand primer (for amino acid substitutions) or with deletion boundaries indicated with a dash. Oligonucleotide primers for mutant detection are shown with 3' mismatched bases underlined.

45°C before being re-tested by PCR using wild-type and mutant primers. Finally, the presence of the mutation was confirmed by sequencing of the relevant part of the *mcm* ORF amplified by PCR from genomic DNA (see Supplementary Information for DNA sequence traces).

RESULTS

HALOARCHAEAL MCM PROTEINS

Prior to embarking on reverse genetic analysis of *Hfx. volcanii* MCM protein function, we undertook a detailed bioinformatic analysis of MCM protein distribution and conservation across the haloarchaea. To identify haloarchaeal MCM proteins, we searched the UniProtKB sequence database using BLAST with the *Hfx. volcanii* MCM protein (HVO_0220, UniProtKB accession number D4GZG5) as the query sequence. Almost 200 proteins were identified in ~120 species belonging to 26 different genera of the *Halobacteriaceae*. To simplify further analysis, we selected a single species as a representative of each genus. These 26 species encoded a total of 39 MCM proteins. A complete list of the species under investigation, together with accession numbers for the proteins we identified, can be found in **Table 1**.

To investigate the relationship between the proteins in greater detail, we constructed multiple protein sequence alignments using ClustalX 2.1 (Larkin et al., 2007). Previously, when analysing a significantly smaller number of genomes (five), we identified a core group of five MCM proteins and a further three proteins that we designated as outliers (MacNeill, 2009). We performed a similar analysis of the larger dataset, using ClustalX to generate multiple sequence alignments of protein sequences (from which inteins were first removed—see below) and njplot (Perriere and Gouy, 1996) to generate unrooted phylogenetic trees. **Figure 1** shows the phylogenetic tree for the 39 proteins. As before, two groups are apparent, corresponding to the core and outlier groups defined previously (MacNeill, 2009). Each of the 26 species encodes a single member of the core group that includes the *Hfx. volcanii* MCM protein. The core group MCM proteins range in length from 697 to 702 amino acids (after inteins are removed—see below) and display a minimum pairwise protein sequence identity of 68% (range 68–95%). Three hundred and twenty four residues (46% of the protein sequence) are absolutely conserved across all 26 core group proteins. This includes the key catalytic residues that make up the Walker A (P-loop) motif (GDPGTGKS in all 26 proteins compared to the consensus GX₄GKS/T), the Walker B motif (DELD in all 26 proteins) and arginine finger motif (SRF in all 26 proteins).

The outlier group proteins display much greater variety in length (312–717 amino acids) and sequence similarity (24–48% identity with *Hfx. volcanii* MCM). The four shortest proteins (312–315 amino acids) are made up of sequences related to the non-catalytic N-terminal domain of MCM only and cannot possess helicase activity. The remaining nine outliers all possess conserved Walker B (DEL/ID) and arginine finger (SRF) motifs, as well as the key lysine in the Walker A (P-loop) motif, suggesting that these proteins may well have ATPase and/or helicase activities. To date, none has been characterized biochemically.

Inteins are parasitic genetic elements capable of efficient self-splicing at the protein level (Gogarten and Hilario, 2006) and are

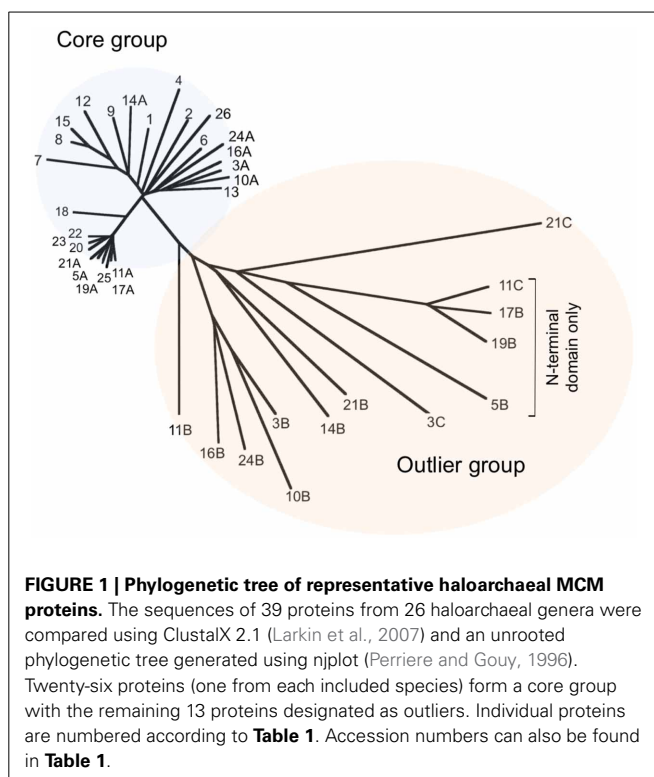
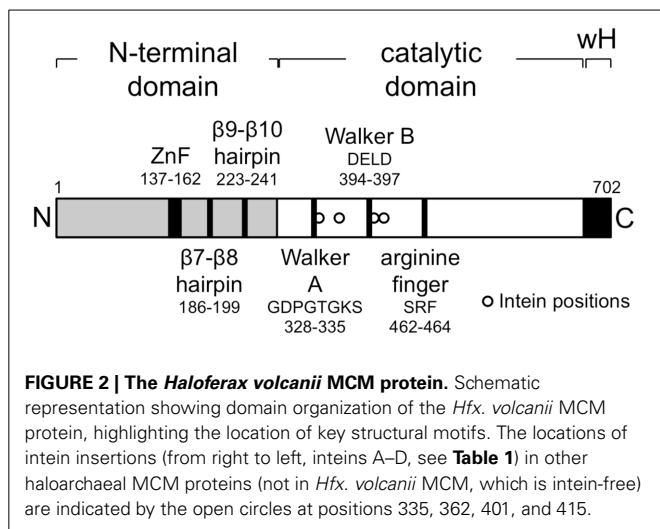


FIGURE 1 | Phylogenetic tree of representative haloarchaeal MCM proteins. The sequences of 39 proteins from 26 haloarchaeal genera were compared using ClustalX 2.1 (Larkin et al., 2007) and an unrooted phylogenetic tree generated using njplot (Perriere and Gouy, 1996). Twenty-six proteins (one from each included species) form a core group with the remaining 13 proteins designated as outliers. Individual proteins are numbered according to **Table 1**. Accession numbers can also be found in **Table 1**.

a common feature of the haloarchaeal MCM proteins. Amongst the 39 proteins, inteins were found in 13, all members of the core group defined above, and at four different locations within the C-terminal catalytic domain of the protein, with the largest number of inteins in an individual MCM protein being four (**Table 1**). **Figure 2** shows the position of the four inteins relative to conserved sequence features found in archaeal MCM proteins. As noted previously, inteins are frequently located at or very near to highly conserved and functionally important sequence regions (Gogarten and Hilario, 2006) and the haloarchaeal MCM inteins are no exception: intein insertion site A is located immediately C-terminal to the essential lysine in the Walker A sequence GDPGTGKS mentioned above, while intein C lies just four amino acids C-terminal to the Walker B sequence DELD (**Table 1**, **Figure 2**). Intein splicing is therefore likely to be essential for protein function.

In a number of archaeal organisms, including representatives of the crenarchaea, euryarchaea, thaumarchaea and korarchaea, the gene encoding MCM is found adjacent to that encoding a GINS subunit (MacNeill, 2010). To ask whether this arrangement is also found in any of the species under investigation in this report, we examined the chromosomal context of each of the 26 core MCM genes and also the 13 outlier proteins. None of the genes encoding core group proteins is located adjacent to a gene encoding GINS or indeed, to any known replication gene (data not shown). A similar situation is seen with genes encoding 12 of the 13 outlier proteins, the sole exception being the pNG3053 protein (UniProtKB accession number Q5V814, labeled 3C in **Figure 1**) encoded by plasmid pNG300 of *Haloarcula marismortui* which is located immediately 3' to an ORF encoding



the C-terminally truncated GINS homolog pNG3052 (data not shown).

REVERSE GENETIC ANALYSIS OF MCM FUNCTION

Hfx. volcanii encodes a single intein-free MCM protein of 702 amino acids in length (**Table 1**) that comprises an N-terminal domain that spans residues 1–283, an AAA+ catalytic core domain spanning residues 283–632 and a C-terminal winged helix-turn-helix (wHTH) domain spanning residues 633–702. All the conserved sequence features characteristic of archaeal MCM proteins are found in the *Hfx. volcanii* protein (**Figure 2**). In order to test whether *Hfx. volcanii* was a workable model for reverse genetic analysis of archaeal MCM function, three of these conserved features were targeted for mutagenesis: the β7–β8 β-hairpin loop (also known as the allosteric communication loop, ACL), the β9–β10 β-hairpin loop and the putative zinc binding domain (**Figure 2**).

To introduce mutations into the *mcm* gene, we used the pop-in/pop-out method (Bitan-Banin et al., 2003). To achieve this, a plasmid was constructed carrying a 1.0 kb region of the *Hfx. volcanii* genome spanning the first 900 nucleotides of the *mcm* open reading frame together with 100 nucleotides of 5′ flanking region. The plasmid also carries the *pyrE2* selectable marker (*pyrE2* function is required for uracil prototrophy) but does not possess a replication origin capable of promoting autonomous replication in *Hfx. volcanii*. Stable maintenance therefore requires that the plasmid integrates (pops-in) into the *Hfx. volcanii* genome by means of homologous recombination between the *mcm* sequences on the plasmid and the *mcm* gene in the chromosome. PCR overlap extension mutagenesis was used to generate a series of mutated forms of the plasmid in which the targeted amino acids were either replaced, singly or in pairs, with one or two alanine residues respectively, or deleted altogether. Next, the plasmids were introduced into *Hfx. volcanii* Δ*pyrE2* strain H26 by standard methods and transformant (pop-in) colonies obtained on Hv-CA plates lacking uracil. Multiple independent colonies were then individually picked and grown in non-selective (uracil-containing) Hv-YPC medium before

being plated onto Hv-CA plates containing 5-fluoroorotic acid to select for (pop-out) clones that had lost the plasmid (see Materials and Methods). Pop-out colonies were then screened by PCR using oligonucleotide primers specific for either the wild-type or mutant sequences (see **Table 3** for primer sequences). Candidate mutant strains were sequentially re-streaked twice on non-selective medium (Hv-YPC) before the presence of the mutation was confirmed by PCR using wild-type- and mutant-specific primers and by sequencing of amplified genomic DNA (see Materials and Methods). **Table 2** lists the strains constructed by this method.

MUTAGENESIS OF THE β7–β8 β-HAIRPIN LOOP

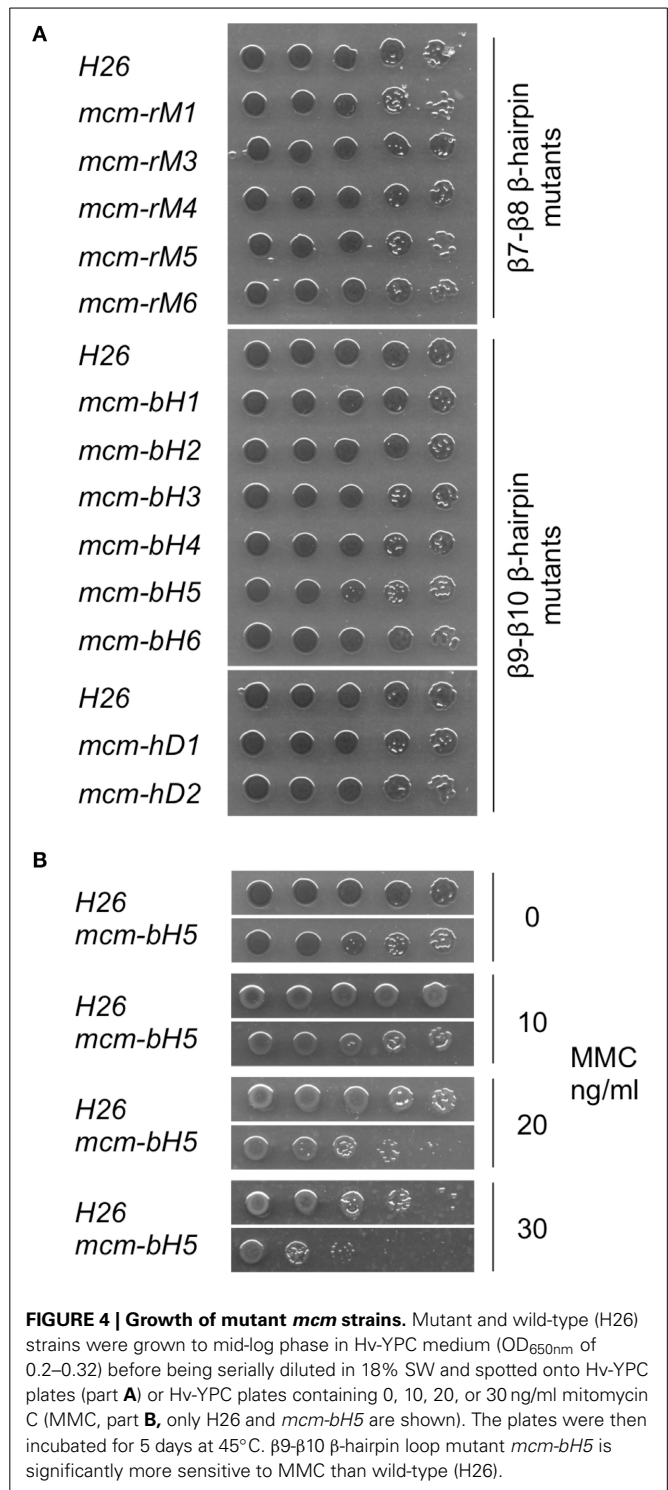
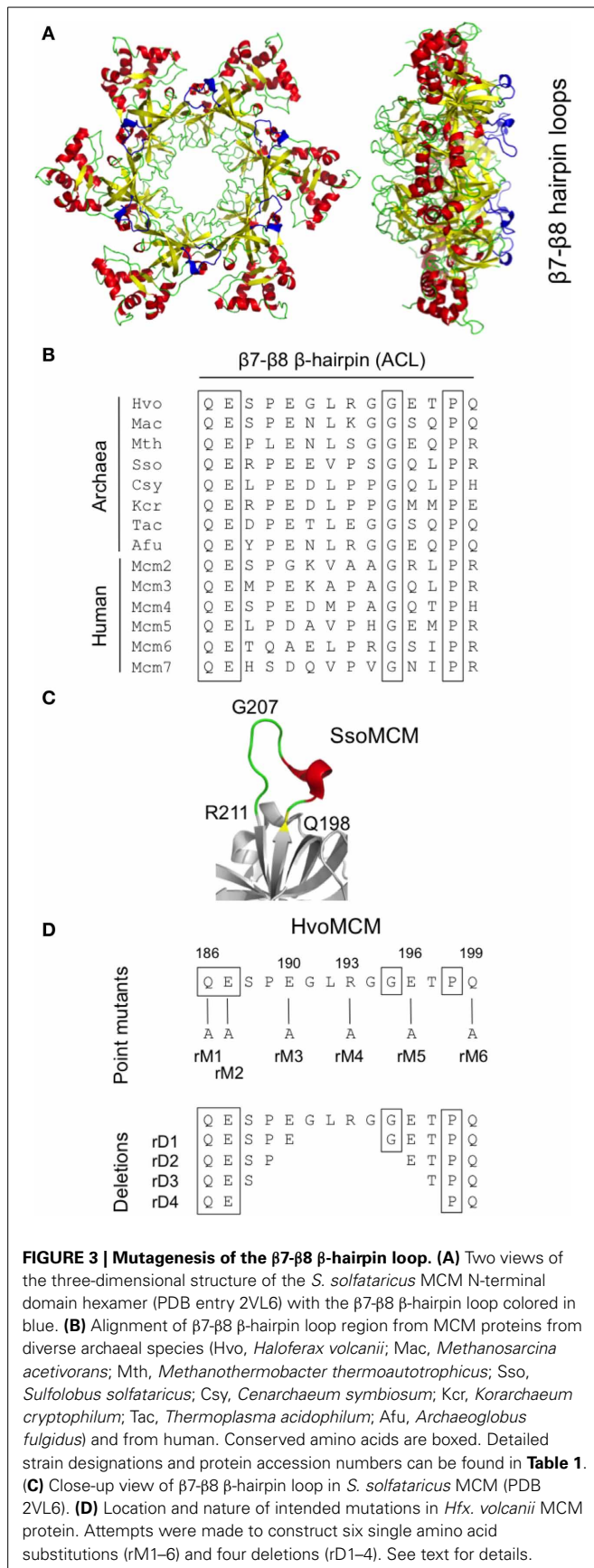
First identified on the basis of its high degree of sequence conservation across species, the β7–β8 β-hairpin loop (also known as the allosteric communication loop, ACL) is located at the interface between the N-terminal and AAA+ catalytic domains of the MCM protein (**Figure 3**). A number of mutants in β7–β8 loop have previously been analyzed biochemically, leading to the conclusion that the loop plays a role in coupling the activities of these two domains (Sakakibara et al., 2008; Barry et al., 2009). The mutations tested include six single point mutants in recombinant *M. thermoautotrophicum* MCM (including four at the boxed conserved residues shown in **Figure 3B**) as well as a triple point mutant and a complete replacement of the loop by the tripeptide serine-asparagine-glycine in *S. solfataricus* MCM (Sakakibara et al., 2008; Barry et al., 2009).

In order to test whether the function of the β7–β8 loop was essential *in vivo* in *Hfx. volcanii*, we attempted to construct 10 different alleles (**Figure 3**): six single point mutants in which individual charged or bulky polar amino acids within the loop were replaced with alanine (mutants *mcm-rM1*–*mcm-rM6*) and four deletions, of two, four, six, and eight amino acids (mutants *mcm-rD1*–*mcm-rD4*, respectively). Two of the individual amino acids targeted (Q186 and E187) are conserved across species (**Figure 3B**); the equivalent residues were mutated in *M. thermoautotrophicum* MCM (Sakakibara et al., 2008).

Of the 10 mutants, five (*mcm-rM1*, *mcm-rM3*–*mcm-rM6*) were readily identified by PCR screening using genomic DNA templates and oligonucleotide primers specific for the wild-type and mutant sequences (see **Table 3**) and confirmed by DNA sequencing (**Supplementary Figure S1**). Despite extensive screening (see Materials and methods), the remaining four strains (*mcm-rM2*, *mcm-rD1*–*mcm-rD4*) could not be isolated, suggesting that these mutations either inactivate or significantly impair the function of the *Hfx. volcanii* MCM protein in cells grown at 45°C. Growth of the five viable mutants was indistinguishable from the parental wild-type strain H26 (**Figure 4A**, upper panel). These results indicate that the β7–β8 loop is indeed essential for *Hfx. volcanii* MCM function although it is also clear that some point mutations (including replacement of conserved amino acid glutamine 186 with alanine) can be tolerated.

MUTAGENESIS OF THE β9–β10 β-HAIRPIN LOOP

Structural analysis of archaeal MCM proteins identifies four β-hairpins in each monomer, three of which protrude, to a greater or lesser extent, into the central channel through which single-



or double-stranded DNA is thought to pass (Slaymaker and Chen, 2012). One of these hairpins, the positively charged β9-β10 β-hairpin (also known as the NT-hairpin) is located in the N-terminal domain of the protein (see Figure 5A). Unlike the β7-β8 loop described above, the sequence of this part of the MCM protein is not well-conserved across evolution (Figure 5B).

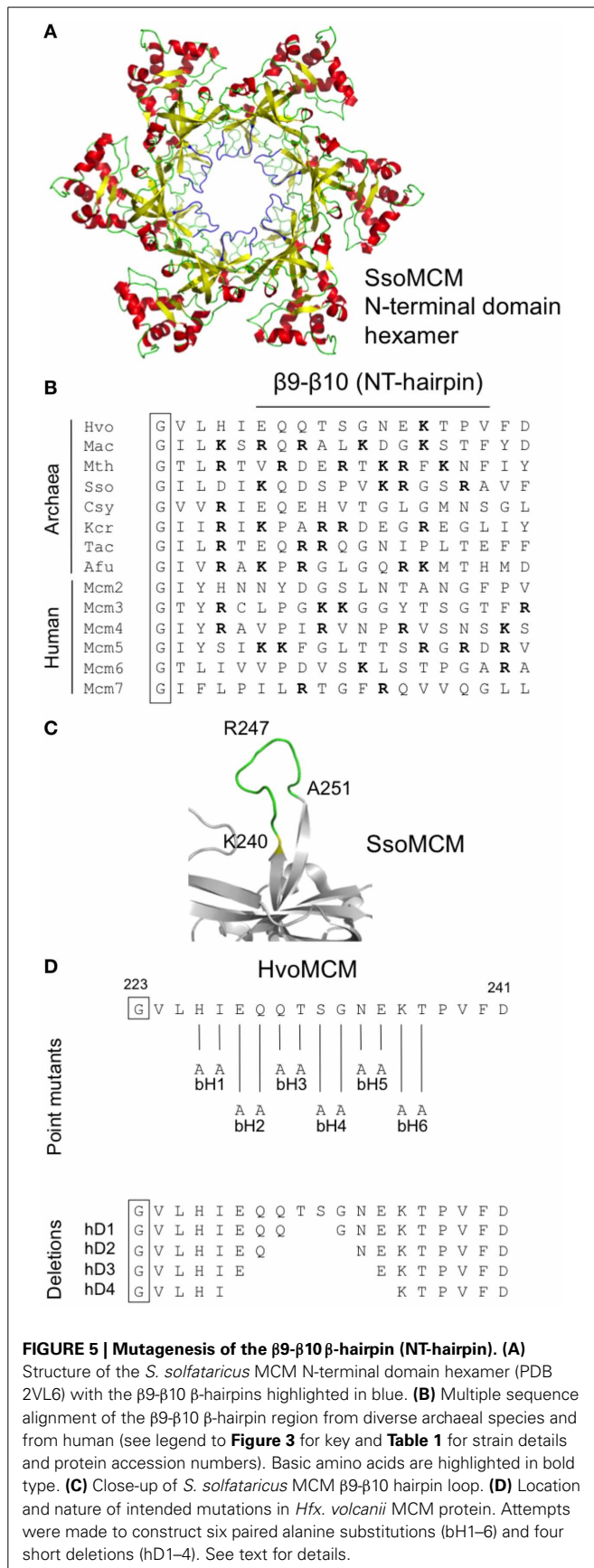


FIGURE 5 | Mutagenesis of the $\beta 9\text{-}\beta 10$ β -hairpin (NT-hairpin). (A)

Structure of the *S. solfataricus* MCM N-terminal domain hexamer (PDB 2VL6) with the $\beta 9\text{-}\beta 10$ β -hairpins highlighted in blue. (B) Multiple sequence alignment of the $\beta 9\text{-}\beta 10$ β -hairpin region from diverse archaeal species and from human (see legend to Figure 3 for key and Table 1 for strain details and protein accession numbers). Basic amino acids are highlighted in bold type. (C) Close-up of *S. solfataricus* MCM $\beta 9\text{-}\beta 10$ hairpin loop. (D) Location and nature of intended mutations in *Hfx. volcanii* MCM protein. Attempts were made to construct six paired alanine substitutions (bH1–6) and four short deletions (hD1–4). See text for details.

However, it appears that the positively charged nature of the hairpin is important for function: mutation of arginine 226 and lysine 228 in the *M. thermoautotrophicum* MCM abolishes that protein's ability to bind to DNA (Fletcher et al., 2003).

To probe the *in vivo* function of the $\beta 9\text{-}\beta 10$ β -hairpin loop in *Hfx. volcanii*, and in the absence of highly conserved amino acids presenting themselves as obvious targets for mutagenesis, we initially attempted to construct six mutants, *mcm-bH1*—*mcm-bH6*, in which adjacent amino acids were replaced with paired alanines (Figure 5D). All six mutants were recovered following PCR screening (and confirmed by sequencing, see Supplementary Figure S2), implying that the $\beta 9\text{-}\beta 10$ β -hairpin loop is readily mutable. None of six mutants exhibited obvious growth defects at 45°C (Figure 4A, middle panel).

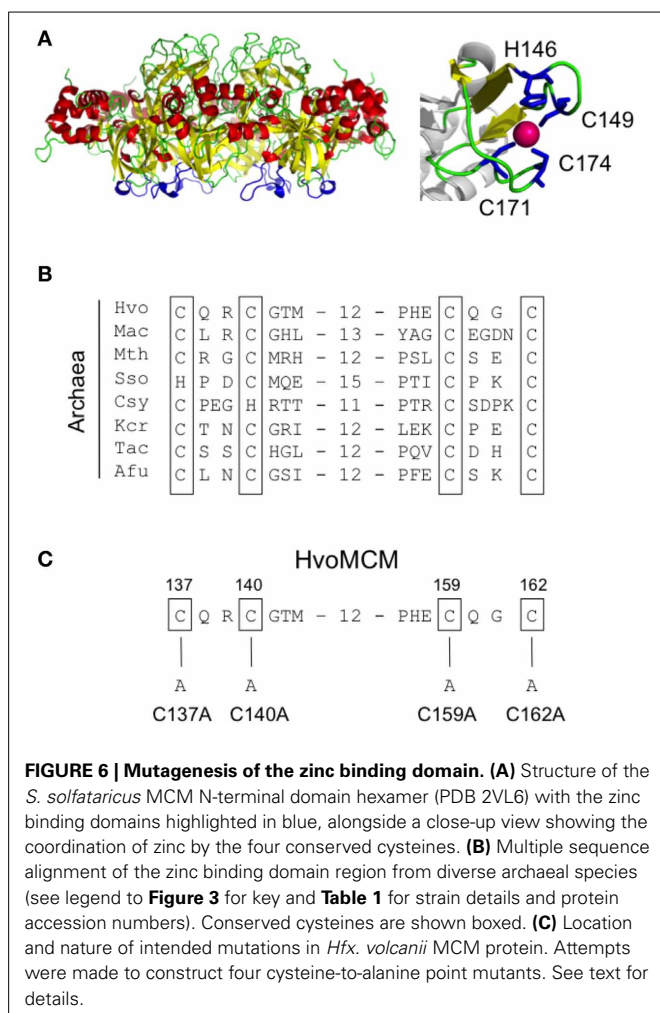
We therefore extended this analysis by attempting to create strains carrying deletions of increasing size in the $\beta 9\text{-}\beta 10$ loop (mutants *mcm-hD1*—*mcm-hD4*, see Figure 5D). However, despite extensive screening, only two of the four could be isolated: *mcm-hD1* and *mcm-hD2*. Growth of these strains, like *mcm-bH1*—*mcm-bH6*, was indistinguishable from the parental wild-type H26 (Figure 4A, lower panel). That we were unable to isolate *mcm-rD3* and *mcm-rD4* strains strongly implies that these deletions either inactivate or significantly impair the function of the *Hfx. volcanii* MCM protein. We conclude from this analysis that while the precise sequence of the $\beta 9\text{-}\beta 10$ β -hairpin loop is not absolutely required for the function of the protein, the loop itself does have a crucial role.

MUTAGENESIS OF THE ZINC BINDING DOMAIN

The N-terminal domain of the archaeal MCM proteins contains cysteine and histidine residues that fold into a zinc binding domain (Figure 6) (Slaymaker and Chen, 2012). A similar domain appears also to be present in eukaryotic MCM proteins but its precise function in either kingdom is unknown. We attempted to construct mutants in which each of the four cysteines was individually replaced with alanine (Figure 6C). However, none of these four mutants that we attempted to construct (mutants *mcm-C137A*, *mcm-C140A*, *mcm-C159A*, and *mcm-C162A*, see Figure 6C) could be isolated, implying that all four cysteines—and presumably zinc binding—is essential for MCM function *in vivo*. As a control, we tested whether it was possible to introduce a silent mutation into the vicinity of the cysteine 140 codon by changing the sequence GGGACG encoding glycine 141 and threonine 142 to an *Acc65I* restriction site, GGTACC (mutation *mcm-S1*). Sixty colonies were screened by *Acc65I* digestion of the PCR amplified *mcm* gene: 26 contained the *Acc65I* site (data not shown), indicating that the region of the gene encoding the zinc binding domain can be mutated in this manner.

SENSITIVITY TO MITOMYCIN C

In total, we isolated 13 mutant strains carrying either single alanine substitutions, paired alanine substitutions and short sequence deletions in two conserved sequence elements, the $\beta 7\text{-}\beta 8$ and $\beta 9\text{-}\beta 10$ β -hairpin loops. None of these strains displayed obvious growth deficiencies: none of the strains grew slowly (Figure 4A) nor were any of the strains cold-sensitive or



temperature-sensitive (data not shown). We also tested whether the strains might be sensitive to DNA damage induced by the DNA modifying agent mitomycin C (MMC). MMC forms three types of MMC-DNA adducts: monoadducts, intrastrand biadducts and interstrand crosslinks (Tomasz, 1995; Bargonetti et al., 2010). Of the 13 tested strains, one strain, *mcm-bH5*, showed increased sensitivity to MMC treatment (**Figure 4B**). This mutant has two adjacent residues in the $\beta 9$ - $\beta 10$ β -hairpin loop, asparagine 234 and glutamate 235, replaced with alanine.

DISCUSSION

The MCM helicase is the key catalytic engine of DNA unwinding during chromosome replication in eukaryotes and most likely in archaea also. Three factors make the archaeal MCM proteins excellent models for their human counterparts. First, the relatively high level of sequence similarity between the eukaryotic and archaeal MCM proteins throughout the non-catalytic N-terminal and catalytic AAA+ domains. Second, the relative simplicity of the homohexameric archaeal MCM complexes compared to the heterohexameric eukaryotic complexes. Third, the relative ease with which certain archaeal MCM proteins can be purified in recombinant form, assayed for various activities and crystallized

for structure determination. It is striking that although more than 10 years have passed since publication of the first partial crystal structure of an archaeal MCM (Fletcher et al., 2003), no high-resolution eukaryotic MCM structures have been solved. Until this occurs, the archaea will remain an important model allowing detailed dissection of MCM protein function.

Using multiple protein sequence alignments and crystal structures as a guide, a number of groups have identified regions of the MCM protein with potentially important roles in MCM function. These have then been mutated and the consequences for MCM activity determined by a variety of *in vitro* biochemical assays. The $\beta 7$ - $\beta 8$ β -hairpin loop was first identified in this way, for example (Sakakibara et al., 2008; Barry et al., 2009). However, despite the growing number of similar studies in the literature, no attempt has been made to examine the *in vivo* consequences of such mutations.

Here we describe the results of probing the *in vivo* function of specific amino acids within an archaeal MCM protein using the haloarchaeon *Hfx. volcanii* as a model. *Hfx. volcanii* encodes a single MCM protein of 702 amino acids. This protein is part of the core group of haloarchaeal MCM proteins defined in a previous study (MacNeill, 2009) and expanded upon here. Each of 26 haloarchaeal species investigated encodes a single member of this group (**Table 1**, **Figure 1**). Ten species encode additional MCM family proteins classed as outliers. Given that a single core protein is found in all species examined and that 16 of the species encode only a single MCM, it is highly likely that these proteins act as replicative helicases. In support of this, the single MCM protein encoded by *Hbt. salinarum* NRC-1 (equivalent to the *Hbt. salinarum* R1 protein listed in **Table 1** and labeled as number 4 in **Figure 1**) has previously been shown to be essential for cell survival (Berquist et al., 2007). While we have not attempted to delete the *Hfx. volcanii mcm* gene in its entirety, our inability to isolate certain mutant *mcm* alleles in this study strongly points to the *Hfx. volcanii mcm* being essential also.

The cellular functions of the outlying haloarchaeal MCM proteins are unknown. With the exception of the group of four C-terminally truncated MCM proteins highlighted in **Figure 1** (proteins labeled 5B, 11C, 17B, and 19B), all the outliers possess intact Walker A, Walker B and arginine finger motifs, suggesting that they may be active as ATPase and/or DNA helicases. Interestingly, *T. kodakarensis* encodes three MCM proteins, all three of which have helicase activity, although only one is essential for cell viability and which therefore is the likely replicative helicase (Ishino et al., 2011; Pan et al., 2011).

In order to determine whether *Hfx. volcanii* presented a workable model for reverse genetic analysis of archaeal MCM function, we targeted three regions of the proteins for investigation. Initially, we focused on the $\beta 7$ - $\beta 8$ β -hairpin loop located at the interface between the N-terminal and catalytic domains (**Figure 3A**) and previously identified as having a role in communicating conformational changes from the DNA binding N-terminal domain to the catalytic AAA+ domain (Sakakibara et al., 2008; Barry et al., 2009). The $\beta 7$ - $\beta 8$ loop is well conserved across eukaryotic and archaeal evolution (**Figure 3B**). In this study, we constructed five strains (*mcm-rM1*, *mcm-rM3*—*mcm-rM6*)

carrying single point mutations in the *Hfx. volcanii* MCM β 7– β 8 β -hairpin loop (Table 2, Figures 3, 4). The mutated residues included glutamine 186, which is conserved in both archaeal and eukaryotic MCM proteins (Figure 3B) but which can be replaced by alanine without significantly affecting cell growth, as well as three charged residues (glutamates 190 and 196, and arginine 193) and glutamine 199, all of which were also replaced by alanine without markedly affecting growth rates. In addition, none of the five β 7– β 8 loop mutants led to increased sensitivity to MMC exposure in spotting assays (Figure 4).

In contrast, we were unable to isolate mutant *mcm-rM2* encoding a protein in which conserved residue glutamate 187 was replaced by alanine, nor any of the four deletion alleles *mcm-rD1*–*mcm-rD4*. Replacement of glutamate 182 in the *M. thermoautotrophicum* MCM protein, the residue corresponding to glutamate 187 in *Hfx. volcanii*, with arginine led to marked reductions in ATPase and helicase activities *in vitro*, without affecting either ATP or DNA binding (Sakakibara et al., 2008), whereas mutation of conserved glutamine 181 (equivalent to glutamine 186 in *Hfx. volcanii* MCM, the residue mutated in *mcm-rM1*) to alanine had little impact. Clearly, conservation is no predictor of *in vivo* essentiality.

We turned next to the β 7– β 8 β -hairpin loop, also known as the NT-hairpin. This loop extends into the central channel of the MCM hexamer and may have a role in tracking DNA through the channel based on the fact that mutating two positively charged residues at the tip of the loop in the *M. thermoautotrophicum* MCM protein abolishes DNA binding *in vitro* (Fletcher et al., 2003). In *H. volcanii*, only a single basic residue is present in the β 9– β 10 loop region, arginine 236. This is likely to lie at or near the tip of the β -hairpin (Figures 5B,C). By comparison with *M. thermoautotrophicum*, it would be reasonable to predict that this amino acid would be essential for N-terminal domain DNA binding and thus for MCM function *in vivo*. In the absence of more widespread sequence conservation, we chose to probe the *in vivo* function of the β 7– β 8 β -hairpin loop initially by attempting to construct a series of six mutants in which adjacent amino acids were replaced with pairs of alanines (Figure 5D). All six mutant strains (Table 2, Figures 3, 5) were viable, including *mcm-bH6* in which arginine 236 was replaced with alanine. Next, we attempted to construct four different β 9– β 10 loop deletion mutants but only two were viable: *mcm-hD1* and *mcm-hD2* (Table 2, Figures 3, 5). The former removes amino acids 231 and 232, and the latter amino acids 230–233. In contrast, strains carrying two larger deletions, *mcm-hD3* and *mcm-hD4*, removing amino acids 229–234 and 228–235, respectively, could not be isolated. Thus, these results clearly demonstrate that while no individual amino acid within the β 9– β 10 hairpin loop is essential for *Hfx. volcanii* MCM function *in vivo*, the loop does have a key role to play.

In addition to mutagenizing the β 7– β 8 and β 9– β 10 loops, we also attempted to individually replace with alanine each of the four cysteine residues that make up the zinc binding domain (Figure 6). However, we were unable to recover any of four desired mutants, *mcm-C137A*, *mcm-C140A*, *mcm-C159A*, or *mcm-C162A*, and conclude that the zinc binding domain is therefore likely to have an essential function *in vivo*. Consistent with

this, replacing the equivalent of *Hfx. volcanii* MCM cysteine 162 with serine produces an *M. thermoautotrophicum* MCM protein with impaired ATPase and single-stranded DNA binding activities and no helicase activity (Poplawski et al., 2001).

Finally, the mutant *Hfx. volcanii* strains generated in this study were tested for increased sensitivity to the DNA modifying drug mitomycin C (MMC). MMC is a potent DNA inter-strand crosslinker and is widely used as replication fork blocking agent, as replication cannot continue past such crosslinks. We tested all 13 viable *mcm* alleles for MMC sensitivity by spotting serially diluted cultures onto medium containing increasing concentrations of MMC and found one strain, *mcm-bH5*, that displayed significantly enhanced sensitivity compared with the parental wild-type H26 (Figure 4B). On medium lacking MMC, growth of *mcm-bH5*, as with the other viable β 9– β 10 loop mutants, is indistinguishable from wild-type (Figure 4A). The *mcm-bH5* protein carries a double substitution in β 9– β 10 loop with asparagine 234 and glutamate 235 being replaced with a pair of significantly less bulky alanines. In the absence of a crystal structure of the *Hfx. volcanii* MCM protein, it is difficult to predict whether the impact of the *mcm-bH5* mutations is confined to the β 9– β 10 loop alone or whether these amino acid changes will have a wider impact on N-terminal domain structure. It is also difficult to envisage how the *mcm-bH5* mutations would cause cells to become supersensitive to MMC, particularly as interpreting the effect of MMC on cells is complicated by the different modes of action of this compound. MMC forms at least four different types of DNA adduct: two species of MMC-mono-dG-adduct, intra-strand dG-MMC-dG biadducts and inter-strand dG-MMC-dG crosslinks, with the latter being assumed to be the replication fork blocking lesion (Tomasz, 1995; Bargonetti et al., 2010). Given the key role of the MCM helicase at the fork it is tempting to speculate that the supersensitivity of *mcm-bH5* is a result of difficulties that arise when the mutant helicase encounters an inter-strand MMC crosslink. However, confirmation of this clearly requires additional experimentation.

In conclusion, the results presented here demonstrate the feasibility of using *Hfx. volcanii* as a model system for reverse genetic analysis of archaeal MCM protein function, provide important confirmation of the *in vivo* importance of conserved structural features identified by previous bioinformatic, biochemical and structural studies, and offer the prospect of more extensive mutational analysis, not only of MCM but of other key replication factors, in the future.

AUTHOR CONTRIBUTIONS

Tatjana P. Kristensen, Reeya Maria Cherian, and Fiona C. Gray performed the bulk of the experimental work. Stuart A. MacNeill designed the study, carried out some of the experimental work and wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank our colleagues in both Copenhagen and St. Andrews for their help with this study. This work was funded by Forskningsrådet for Natur og Univers (FNU sagsnr. 272-05-0446).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00123/abstract>

Supplementary Figure S1 | Sequence analysis of $\beta 7$ - $\beta 8$ β -hairpin mutant strains. Codons mutated to encode alanine are boxed.

Supplementary Figure S2 | Sequence analysis of $\beta 9$ - $\beta 10$ β -hairpin mutant strains. Codons mutated to encode alanine pairs (mutants bH1–6) are boxed. Vertical lines in hD1 and hD2 indicate location of deleted sequences.

REFERENCES

- Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Bae, B., Chen, Y. H., Costa, A., Onesti, S., Brunzelle, J. S., Lin, Y., et al. (2009). Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure* 17, 211–222. doi: 10.1016/j.str.2008.11.010
- Bargonetti, J., Champeil, E., and Tomasz, M. (2010). Differential toxicity of DNA adducts of mitomycin C. *J. Nucleic Acids* 2010:698960. doi: 10.4061/2010/698960
- Barry, E. R., Lovett, J. E., Costa, A., Lea, S. M., and Bell, S. D. (2009). Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1051–1056. doi: 10.1073/pnas.0809192106
- Berquist, B. R., Dassarma, P., and Dassarma, S. (2007). Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. *BMC Genet* 8:31. doi: 10.1186/1471-2156-8-31
- Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J. Bacteriol.* 185, 772–778. doi: 10.1128/JB.185.3.772-778.2003
- Brewster, A. S., Wang, G., Yu, X., Greenleaf, W. B., Carazo, J. M., Tjajadi, M., et al. (2008). Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20191–20196. doi: 10.1073/pnas.0808037105
- Duderstadt, K. E., and Berger, J. M. (2008). AAA+ ATPases in the initiation of DNA replication. *Crit. Rev. Biochem. Mol. Biol.* 43, 163–187. doi: 10.1080/10409230802058296
- Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., et al. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20240–20245. doi: 10.1073/pnas.0911500106
- Farkas, J. A., Picking, J. W., and Santangelo, T. J. (2013). Genetic techniques for the archaea. *Annu. Rev. Genet.* 47, 539–561. doi: 10.1146/annurev-genet-111212-133225
- Fletcher, R. J., Bishop, B. E., Leon, R. P., Sclafani, R. A., Ogata, C. M., and Chen, X. S. (2003). The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nat. Struct. Biol.* 10, 160–167. doi: 10.1038/nsb893
- Fu, Y., Slaymaker, I. M., Wang, J., Wang, G., and Chen, X. S. (2014). The 1.8-Å crystal structure of the N-terminal domain of an archaeal MCM as a right-handed filament. *J. Mol. Biol.* 426, 1512–1523. doi: 10.1016/j.jmb.2013.12.025
- Gambus, A., Khoudoli, G. A., Jones, R. C., and Blow, J. J. (2011). MCM2-7 form double hexamers at licensed origins in *Xenopus* egg extract. *J. Biol. Chem.* 286, 11855–11864. doi: 10.1074/jbc.M110.199521
- Gogarten, J. P., and Hilario, E. (2006). Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* 6:94. doi: 10.1186/1471-2148-6-94
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature* 503, 544–547. doi: 10.1038/nature12650
- Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., et al. (2011). Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells* 16, 1176–1189. doi: 10.1111/j.1365-2443.2011.01562.x
- Kamada, K. (2012). The GINS complex: structure and function. *Subcell. Biochem.* 62, 135–156. doi: 10.1007/978-94-007-4572-8_8
- Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F. M., and Onesti, S. (2012). Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. *J. Biol. Chem.* 287, 4121–4128. doi: 10.1074/jbc.M111.285395
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., et al. (2011). A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res.* 39, 6114–6123. doi: 10.1093/nar/gkr181
- Liu, W., Pucci, B., Rossi, M., Pisani, F. M., and Ladenstein, R. (2008). Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Res.* 36, 3235–3243. doi: 10.1093/nar/gkn183
- MacNeill, S. A. (2009). The haloarchaeal chromosome replication machinery. *Biochem. Soc. Trans.* 37, 108–113. doi: 10.1042/BST0370108
- MacNeill, S. A. (2010). Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem. J.* 425, 489–500. doi: 10.1042/BJ20091531
- Magrane, M., and UniProt Consortium. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011:bar009. doi: 10.1093/database/bar009
- Makarova, K. S., Koonin, E. V., and Kelman, Z. (2012). The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol. Direct* 7:7. doi: 10.1186/1745-6150-7-7
- Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., and Bell, S. D. (2006). GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep.* 7, 539–545. doi: 10.1038/sj.embor.7400649
- Morgunova, E., Gray, F. C., MacNeill, S. A., and Ladenstein, R. (2009). Structural insights into the adaptation of proliferating cell nuclear antigen (PCNA) from *Haloferax volcanii* to a high-salt environment. *Acta Crystallogr. D Biol. Crystallogr.* 65, 1081–1088. doi: 10.1107/S0907444909029321
- Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H., and Allers, T. (2007). Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet.* 3:e77. doi: 10.1371/journal.pgen.0030077
- Onesti, S., and MacNeill, S. A. (2013). Structure and evolutionary origins of the CMG complex. *Chromosoma* 122, 47–53. doi: 10.1007/s00412-013-0397-x
- Oyama, T., Ishino, S., Fujino, S., Ogino, H., Shirai, T., Mayanagi, K., et al. (2011). Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol.* 9:28. doi: 10.1186/1741-7007-9-28
- Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N., and Kelman, Z. (2011). *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res.* 39, 9671–9680. doi: 10.1093/nar/gkr624
- Perler, F. B. (2002). InBase: the intein database. *Nucleic Acids Res.* 30, 383–384. doi: 10.1093/nar/30.1.383
- Perriere, G., and Gouy, M. (1996). WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* 78, 364–369. doi: 10.1016/0300-9084(96)84768-7
- Poidevin, L., and MacNeill, S. A. (2006). Biochemical characterisation of LigN, an NAD⁺-dependent DNA ligase from the halophilic euryarchaeon *Haloferax volcanii* that displays maximal *in vitro* activity at high salt concentrations. *BMC Mol. Biol.* 7:44. doi: 10.1186/1471-2199-7-44
- Poplawski, A., Grabowski, B., Long, S. E., and Kelman, Z. (2001). The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. *J. Biol. Chem.* 276, 49371–49377. doi: 10.1074/jbc.M108519200
- Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., and Diffley, J. F. (2009). Concerted loading of MCM2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139, 719–730. doi: 10.1016/j.cell.2009.10.015
- Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F. P., and Kelman, Z. (2008). Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Res.* 36, 1309–1320. doi: 10.1093/nar/gkm1160
- Sanchez-Pulido, L., and Ponting, C. P. (2011). Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics* 27, 1885–1888. doi: 10.1093/bioinformatics/btr332
- Skowrya, A., and MacNeill, S. A. (2012). Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res.* 40, 1077–1090. doi: 10.1093/nar/gkr838

- Slaymaker, I. M., and Chen, X. S. (2012). MCM structure and mechanics: what we have learned from archaeal MCM. *Subcell. Biochem.* 62, 89–111. doi: 10.1007/978-94-007-4572-8_6
- Slaymaker, I. M., Fu, Y., Toso, D. B., Ranatunga, N., Brewster, A., Forsburg, S. L., et al. (2013). Mini-chromosome maintenance complexes form a filament to remodel DNA structure and topology. *Nucleic Acids Res.* 41, 3446–3456. doi: 10.1093/nar/gkt022
- Stroud, A., Liddell, S., and Allers, T. (2012). Genetic and biochemical identification of a novel single-stranded DNA-binding complex in *Haloflex volcanii*. *Front. Microbiol.* 3:224. doi: 10.3389/fmicb.2012.00224
- Tomasz, M. (1995). Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* 2, 575–579. doi: 10.1016/1074-5521(95)90120-5
- Vijayraghavan, S., and Schwacha, A. (2012). The eukaryotic Mcm2-7 replicative helicase. *Subcell. Biochem.* 62, 113–134. doi: 10.1007/978-94-007-4572-8_7
- Walters, A. D., and Chong, J. P. (2010). An archaeal order with multiple minichromosome maintenance genes. *Microbiology* 156, 1405–1414. doi: 10.1099/mic.0.036707-0
- Wendoloski, D., Ferrer, C., and Dyal-Smith, M. L. (2001). A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloflex volcanii* strain cured of plasmid pHV2. *Microbiology* 147, 959–964.
- Wiedemann, C., Ohlenschläger, O., Medagli, B., Onesti, S., and Görlach, M. (2013). ¹H, ¹⁵N and ¹³C chemical shift assignments for the winged helix domains of two archaeal MCM C-termini. *Biomol. NMR Assign.* doi: 10.1007/s12104-013-9516-0. [Epub ahead of print].
- Winter, J. A., Christofi, P., Morroll, S., and Bunting, K. A. (2009). The crystal structure of *Haloflex volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct. Biol.* 9:55. doi: 10.1186/1472-6807-9-55
- Yuan, H., Liu, X.-P., Han, Z., Allers, T., Hou, J.-L., and Liu, J.-H. (2013). RecJ-like protein from *Pyrococcus furiosus* has 3'-5' exonuclease activity on RNA: implications for proofreading of 3'-mismatched RNA primers in DNA replication. *Nucleic Acid Res.* 41, 5817–5826. doi: 10.1093/nar/gkt275
- Zhao, A., Gray, F. C., and MacNeill, S. A. (2006). ATP- and NAD⁺-dependent DNA ligases share an essential function in the halophilic archaeon *Haloflex volcanii*. *Mol. Microbiol.* 59, 743–752. doi: 10.1111/j.1365-2958.2005.04975.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 February 2014; accepted: 10 March 2014; published online: 26 March 2014.

Citation: Kristensen TP, Maria Cherian R, Gray FC and MacNeill SA (2014) The haloarchaeal MCM proteins: bioinformatic analysis and targeted mutagenesis of the β 7- β 8 and β 9- β 10 hairpin loops and conserved zinc binding domain cysteines. *Front. Microbiol.* 5:123. doi: 10.3389/fmicb.2014.00123

This article was submitted to *Extreme Microbiology*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Kristensen, Maria Cherian, Gray and MacNeill. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.